COMPARISON OF THE ESTIMATES ON LANGUAGE USE AND ENGLISH-SPEAKING ABILITY FROM THE ACS, THE C2SS, AND CENSUS 2000

Hyon B. Shin

Housing and Household Economics Statistics Division
U.S. Census Bureau

ESTIMATES ABOUT LANGUAGE USE AND ENGLISH-SPEAKING ABILITY FROM THE ACS, THE C2SS, AND CENSUS 2000

INTRODUCTION

This report compares national distributions based on data from the American Community Survey (ACS) with those based on data from various Census Bureau surveys for three items: speaking a language other than English at home, the languages spoken, and English-speaking ability. This report first compares estimates of the number of people who reported speaking a language other than English at home. It then notes variations that are both statistically and substantially different, and for those found, it offers possible explanations.

The second analysis compares specific languages spoken. The 2005 ACS data are compared with Census 2000 sample data and the Census 2000 Supplementary Survey (C2SS) data because the Census 2000 data constitute the most comparable data.

The third analysis compares the English-speaking ability of respondents. The four categories of English-speaking ability are "very well," "well," "not well," and "not at all."

METHODOLOGY

The tables included in this report compare the most commonly tabulated data on language spoken at home and English-speaking ability from the ACS, the C2SS, and

Census 2000. Comparisons consist primarily of percent and percentage-point differences between the ACS and all other comparison distributions. Tables display the survey estimates, the margins of error from which 90-percent confidence intervals of the estimates can be derived, and the difference between the estimates. In the case of frequency distributions, the difference is calculated as the percent difference between the two estimates. In the case of relative frequency distributions, the difference is calculated as the percentage-point difference between the two estimates. An asterisk (*) denotes statistically significant differences.

At the national level, survey variances were small, resulting in many statistically significant differences between the distributions. This report focuses on statistically significant differences of 0.5 or more. This yardstick can vary based on the relative size of the category. For example, for population groups constituting a relatively large percentage of the population (such as the percentage of Spanish speakers), a 0.5 percentage-point difference in the estimates might be relatively small, while for population groups constituting a smaller percentage of the population (such as the percentage of Japanese speakers), a 0.5 percentage-point difference could be relatively large. Users may choose statistically significant differences that are smaller or larger than 0.5 for their own analytical purposes.

The remainder of this section examines differences in methodology between these different data sets.

Sample Frame

The ACS began full implementation in 2005 and long-form data are now collected from a national sample of 3 million households a year. The 2005 ACS surveyed a national sample of housing units, both occupied and vacant. An initial sample of 2,922,656 households resulted in 1,924,527 completed interviews. Data were collected in all counties (3,141 counties) in the United States. The sample is designed to provide estimates of housing and socio-economic characteristics for the nation, all states, most areas with a population of 250,000 or more, and selected areas of 65,000 or more.

The long-form questionnaire used in Census 2000 was sent to a sample of approximately 1-in-6 households. This sample was designed to produce national, state, and substate estimates of many social and economic characteristics from questions that were not included on the Census 2000 short-form that was sent to the entire population.[1]

The C2SS was conducted as part of Census 2000 to demonstrate the operational feasibility of ACS methods. The C2SS distributions in this report come from information collected in the year 2000 from the original 36 ACS test counties plus another sample of 1,203 counties selected and surveyed using current ACS operational and data collection methods.[2]

[1] For a detailed explanation of the Census 2000 sampling frame and data collection procedures, see U.S. Census Bureau, *Census 2000 Summary File 3: Technical Documentation.* U.S. Census Bureau: Washington, DC 2002, Chapter 8.

[2] For a detailed explanation of the C2SS survey and comparisons with Census 2000 sample items, see U.S. Census Bureau, *Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey. Report 9: Comparing Social Characteristics with Census 2000.* Washington, DC, 2004.

Sample Size and Mode of Data Collection

The 2005 ACS interviewed a total of 1,924,527 households. Data were collected continuously throughout the year using a combination of mail-out/mail-back questionnaires, Computer-Assisted Telephone Interviewing (CATI), and Computer-Assisted Personal Interviewing (CAPI). Each month a unique national sample of addresses received an ACS questionnaire. Addresses that did not respond were telephoned during the second month of collection if a phone number for the address was available, and personal visits were conducted during the third month and the last month of data collection for a subsample of the remaining nonresponding units. The 2005 ACS achieved an overall survey response rate of 97.3 percent, calculated as the initially weighted estimate of interviews divided by the initially weighted estimate of cases eligible to be interviewed.

In the Census 2000 Supplementary Survey (C2SS), the following ACS methods were used: mailed questionnaire, telephone, and personal-visit data collection over a rolling three-month time period, collecting data from twelve independent monthly samples of addresses during the year. The combination of the 36 ACS test counties with the 1,203 counties selected and surveyed using ACS methods accounted for a sample of almost 900,000 households used to produce the C2SS estimates in this report. The unit survey-

response rate for the C2SS was 95.1 percent. [3] In Census 2000, the sample households had a unit survey-response rate of 91.2 percent.[4]

Residence Rules

The ACS, the C2SS, and Census 2000 used different residence rules to determine which individuals in a household would be eligible for interview. This difference may contribute to variation in the universes from which the social characteristics are measured.

The ACS and C2SS used similar rules. These surveys collected interviews from everyone in the housing unit on the day of interview who was living or staying there for more than two months, regardless of whether or not they maintained a usual residence elsewhere or if they did not have a usual residence elsewhere. If a person who usually lived in the housing unit was away for more than two months at the time of the survey contact, he or she was not considered to be a current resident of that unit. This rule recognizes that people can have more than one place where they live or stay over the course of a year, and these people affect the estimate of the characteristics of the population for some areas. Because the 2005 ACS and the C2SS excluded group quarters from the sample frame and interviewed individuals at their current residence, college students living in dormitories were not included in the ACS universe.

---

[3] For more information on response rates for the ACS and for C2SS see: U.S. Census Bureau, Using the Data: Quality Measure, http://www.census.gov/acs/www/acs-php/quality_measures_response_2006.php

[4] Deborah Griffin, Susan Love, and Sally Obenski, "Can the American Community Survey Replace the Census Long Form?" Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Nashville, TN, May 14-18, 2003.

For Census 2000, each person was enumerated as an inhabitant of his or her "usual residence." Usual residence is the place where the person lives and sleeps most of the time. If a person had no usual residence, the person was to be counted where he or she was staying on Census Day (April 1, 2000). College students were counted as residents of the area in which they were living while attending college. Children in boarding schools below the college level were counted at their parental home.

Question Wording and Reference Periods

The same set of questions on language use and English-speaking ability were asked in the ACS, the C2SS, and Census 2000.

Language spoken at home and English-speaking ability. The 2005 ACS, the C2SS, and Census 2000 asked the following language questions all persons 5 years and over:

Figure 1.
Reproduction of the Question on Language Use From the ACS, the C2SS, and Census 2000

a. Does this person speak a language other than English at home?

❒ Yes
❒ No

b. What is this language?
❒❒❒❒❒❒❒❒❒❒❒❒❒❒❒❒
(*For example: Korean, Italian, Spanish, Vietnamese)*

c. How well does this person speak English?
❒ Very well
❒ Well
❒ Not well
❒ Not at all

The second question, "What is this language?" was a write-in question. Although linguists recognize several thousand languages in the world, the coding operation used by the Census Bureau put the reported languages in about 380 language categories of single languages or language families.[5] The 380 language categories were then recoded to represent two different ways of presenting the detailed languages. The first was to combine all the languages into four major language groups. The second combination was the list of 20 languages, which gave more detail.

Item Nonresponse

Item nonresponse occurs when an individual does not provide complete and usable information for a data item. Item allocation rates are often used as a measure of the level of item nonresponse. These rates are computed as the ratio of the number of eligible people for which a value was allocated during the editing process for a specific item to the number of people eligible to have responded to that item.

Item Allocation Rate	**2005 ACS**	**2000 C2SS**	**Census 2000**
Spoke another language at home	1.7 percent	4.3 percent	6.2 percent
Other language spoken	4.0 percent	8.9 percent	12.0 percent
English-speaking ability	2.5 percent	6.0 percent	8.2 percent

[5] More detailed information on languages and language coding can be found in "Census 2000 Summary File 3 Technical Documentation/prepared by the U.S. Census Bureau, 2002" (*www.census.gov/prod/cen2000/doc/sf3.pdf*).

Spoke a language other than English at home. For the 2005 ACS, the allocation rate for those who spoke a language other than English at home was 1.7 percent. For the C2SS, the allocation rate was 4.3 percent and for Census 2000, the allocation rate was 6.2 percent. The ACS methods of item allocation, which both the ACS and the C2SS use, reduce item nonresponse by conducting follow-up interviews via telephone and personal visits. Census 2000 allowed proxy responses but the added follow-up used by the ACS method may contribute toward the relatively smaller allocation rates for the 2005 ACS and the C2SS. These procedures may be allocating more other language speakers in Census 2000 compared to the 2005 ACS and the C2SS.

Language spoken. The 2005 ACS had an allocation rate of 4.0 for the individual language spoken in the home. The C2SS allocated 8.9 percent of cases and Census 2000 allocated 12.0 percent. The ACS methods of item allocation may be contributing to the relatively lower allocation rates for the 2005 ACS and the C2SS compared to Census 2000.[6] Since Census 2000 allowed proxy answers, the languages spoken answer may have a higher allocation rate if the proxy answer in the language spoken field was inconsistent with other edit checks.

English-speaking ability. Of those who spoke a language other than English at home, the ACS reported an allocation rate of 2.5 percent for the English-speaking ability question. The C2SS reported an allocation rate of 6.0 percent and Census 2000 reported an allocation rate of 8.2 percent. The different modes of collection among the ACS, the

[6] For more information on modes of collection among the 2005 ACS, the C2SS, and Census 2000 see the section on Sample Size and Modes of Collection.

C2SS, and Census 2000 could be contributing to the relatively lower allocation rates for the 2005 ACS and the C2SS compared to Census 2000. The overall allocation rates were lower for the ACS compared to the C2SS and to Census 2000 that could also contribute to the lower ACS allocation rates.

Data Editing and Imputation Procedures

The ACS, the C2SS, and Census 2000 edit and imputation rules are designed to ensure that the final edited data are as consistent and complete as possible. These rules are used to identify and account for missing, incomplete, and contradictory responses. In each case where a problem is detected, pre-established edit rules govern its resolution.

The three surveys employ two principal imputation methods: relational imputation and hot-deck allocation. Relational imputation assigns values for blank or inconsistent responses on the basis of other characteristics on the person's record or within the household. Hot-deck allocation supplies responses for missing or inconsistent data from similar responding housing units or people in the sample.

The editing procedures for all surveys employ logical checking routines to produce consistency among household members and other responses. For example, a person under age 5 should not have any language-related characteristic. When answers cannot be logically assigned or when inconsistencies or missing data are encountered, allocation routines using hot decks generally stratify the donors and recipients of the hot deck by their age, sex, race, and other characteristics of the household.

Controls and Weighting

There are notable differences among the surveys in the selection of controls and the calculation of weights that may lead to differences in estimates. The ACS, the C2SS, and Census 2000 samples are weighted to account for both the probability of selection and housing unit nonresponse. The ACS, the C2SS, and Census 2000 samples are weighted to account for both the probability of selection and housing unit nonresponse.

Data from the C2SS included a series of weighting adjustments. The first factor (initial weight) is the multiplicative inverse of the unbiased probability of selecting the address. Another factor adjusts for the subsampling of nonresponding units before CAPI interview. Another adjusts for housing unit nonresponse. In addition, the survey was controlled at the county level to census counts of the population of individuals and housing units, resulting in both person weights and housing-unit weights. This "weighting to population control totals" adjusts for potential survey undercoverage and compensates for errors not corrected by the other weighting techniques.[7]

The 2005 ACS data included a number of similar weighting adjustments with several differences. Differences between the C2SS and the 2005 ACS include the size of the surveys, and the date of survey controls. The C2SS had sample in selected counties of the US, while the 2005 ACS had sample in every county of the U.S. The C2SS initial weight included the probabilities of selecting counties, as well as the probabilities of

[7] For more information on the C2SS sample design and weighting see: U.S. Census Bureau, Accuracy of the Data (2000) http://www.census.gov/acs/www/Downloads/ACS/Accuracy00_C2SS.pdf

selecting addresses within county. The 2005 ACS initial weight included only the within-county probabilities. The controls for the C2SS were the total housing units and total persons as of April 2000, while the controls for the 2005 ACS were independent estimates of total housing units and total persons as of July 2005.[8]

Estimates from the Census 2000 sample were obtained from an iterative ratio-estimation procedure that assigned a weight to each sample person. The estimation procedure used to assign the weights was performed in geographically defined weighting areas that were usually formed of contiguous geographic units within counties. Within a weighting area, the long-form sample was ratio-adjusted to equal the 100-percent total for certain categories such as family households or nonfamily households, age, sex, race, and Hispanic origin. This procedure resulted in weights for each person that could vary from person to person within the same housing unit.

RESULTS

The differences described in this report comprise two separate analyses. The C2SS and Census 2000 capture the differences between two different data capture vehicles conducted during the same time period. The 2005 ACS and the C2SS differences are real changes since these two sources collected data in two different time periods.

The results the following sections describe the differences between the 2005 ACS and the C2SS first and then the differences between the C2SS and Census 2000.

[8] For more information of the 2005 ACS sample design and weighting see: U.S. Census Bureau, Accuracy of the Data (2005), http://www.census.gov/acs/www/Downloads/ACS/accuracy2005.pdf More details on the ACS design can be found at: http://www.census.gov/acs/www/Downloads/tp67.pdf.

Language spoken and English-speaking ability.

The 2005 American Community Survey (ACS) estimated that of the U.S. population 5 years and over, 52 million people spoke a language other than English at home (19 percent) and the Census 2000 Supplemental Survey (C2SS) estimated 45 million speakers (17 percent) (see Table 1). Of those who spoke a language other than English at home, the ACS estimated 23 million (45 percent) people speaking English less than "very well" while the C2SS estimated 19 million (43 percent).

Between the 2005 ACS and the C2SS, the percentage difference of those who spoke a language other than English at home was 16.7 percent. For those who spoke English less than "very well", there was a 20.1 percent difference. Overall, the percentage differences between the 2005 ACS and the C2SS are most likely due to the increase of non-English language speakers between 2000 and 2005.

While all four language groups had an increase in the number of speakers between the 2005 ACS and the 2000 C2SS, the smallest increase in the number of speakers was for those who spoke Other Indo-European languages (5.6 percent change). The number of Spanish speakers increased by 20.8 percent, Asian and Pacific Island languages increased by 14.8 percent, and all other languages increased by 20.1 percent.[9] Additionally, the number of people aged 5 years or over who spoke a language other than English at home and spoke English less than "very well" increased between the two surveys.

[9] The increase in the number of Spanish speakers (20.8 percent) was not statistically different from the increase in the number of all other language speakers (20.1 percent).

The Census 2000 Supplemental Survey (C2SS) estimated 45 million speakers (17 percent) and Census 2000 estimated 46 million speakers of a non-English language (18 percent). The C2SS estimated 19 million (43 percent) people and Census 2000 estimated 21 million (46 percent) people speaking English less than "very well."

A higher percentage of those who speak English less than "very well" were captured in Census 2000 compared to the C2SS. The difference in the number of people who spoke a language other than English at home between the C2SS and Census 2000 was –2.8 percent, where Census 2000 had a higher estimate. Of those who spoke English less than "very well" however, there was a –8.0 percent difference.

With the exception of those who reported speaking Asian and Pacific Island languages, all percentage differences between the C2SS and Census 2000 were statistically different. Within each of the four major language groups, the percentage difference of those who spoke English less than "very well" was greater than the percentage difference of those who spoke a language other than English at home.

The differences between the C2SS and Census 2000 may be due to different data collection methods, different sample frames, or from different residence rules. A non-methodological reason behind some of these differences could be a result of the Census 2000 advertising campaign. Census 2000 had a massive advertising campaign to bolster the response rates. As part of the campaign, posters and advertisements in different

languages were produced. These language-based advertisements may have contributed to the Census 2000 capturing more non-English language-speaking respondents.

In addition, the Census 2000 questionnaire was printed in five different languages aside from English. These languages were Spanish, Chinese, Tagalog, Vietnamese, and Korean. Census 2000 also had language guides in 49 languages.[10] The ACS prints questionnaires in English and Spanish.

Detailed languages spoken.

The non-English languages most commonly spoken in the U.S. are listed in Table 2. The detailed languages are listed by the number of speakers from the 2005 American Community Survey (ACS). Of the 20 languages listed in Table 2, 15 languages had a statistical difference between the 2005 ACS and the C2SS. Among the twenty languages, Hindi speakers had the highest percentage difference with 45.8 percent. The other languages ranged between having a negative difference of –8.5 percent to 27.0 percent. Some of these differences are probably due to real changes while some may be due to sampling variability due to small sample sizes.

The percentage portion of Table 2 shows the percentage distribution of the twenty languages. Since the vast majority of other language speakers speak Spanish (62.0 percent in the 2005 ACS and 59.8 percent in the C2SS), the remaining languages share the remainder. Of these, only German and French speakers had statistical differences

[10] The list of languages that had language guides in Census 2000 is available at http://www.census.gov/dmd/www/genfaq.htm.

greater than 0.5 between the 2005 ACS and the C2SS (both -0.6 percentage-point difference). This difference is most likely due to a real change in the proportions of German and French speakers among those who spoke a language other than English at home.

The differences between the C2SS and Census 2000 include more variability in what languages have significant differences. Spanish (-3.0 percent), French (-8.2 percent), German (-7.7 percent), Italian (-10.2 percent), and Greek (-15.4 percent) speakers all had negative differences between the C2SS and Census 2000, which means that Census 2000 captured more of these language speakers than in the C2SS.[11]

These differences could be a result of different data collection methods or because of the concerted effort to advertise Census 2000. These differences could also arise from higher allocation rates in Census 2000 where more people who spoke a language other than English at home were allocated (see section on Item Nonresponse).

Language spoken at home for regions and states.

The population 5 years and over who spoke a language other than English at home varied by region and state (see Table 3a). Between the 2005 ACS and C2SS, the South had the greatest gain in the number of people who spoke a language other than English at home (22.6 percent). The Northeast (9.3 percent), the Midwest (13.6 percent), and the West

[11] The percentage change for German speakers (-7.7 percent) was not significantly different from the percentage change of Spanish speakers (-3.0 percent). The percentage change for French (-8.2 percent), German (-7.7 percent), Italian (-10.2 percent), and Greek (-15.4 percent) were not statically different from one another.

(17.2 percent) also had significant gains in their populations speaking a language other than English.[12] In addition, between the two surveys, three out of four states had an increase in the number of people who spoke a language other than English at home.

Among the regions, Census 2000 captured more people who spoke a language other than English at home compared to the C2SS. The percentage difference between the Northeast (-3.5 percent), the Midwest (-1.7 percent), and the South (-3.6 percent) were not significantly different from each other. The West had a percentage difference of –2.1 percent.

Census 2000 also captured more people who spoke a language other than English at home in 48 out of 50 states and the District of Columbia, although only 22 states had significant differences.[13] There's no discernible pattern to show why some states had significant differences while others did not. There were larger states such as California, New York, and Pennsylvania but also smaller states such as Arkansas, Kansas, and South Dakota (see Table 3a).

Examining the percentages of people who spoke a language other than English at home, Table 3b shows that there was a significant percentage-point increase between the 2005 ACS and the C2SS. In all regions and most states, there was an increase in the

[12] Between the 2005 ACS and the C2SS, the percentage change of the population who spoke a language other than English at home in the Midwest (13.6 percent) was not significantly different than the population living in the West (17.2 percent).

[13] The two states in which Census 2000 had a positive percentage change in the number of people 5 years and over who spoke a language other than English at home, when compared to the C2SS were Idaho (2.2 percent) and Indiana (23.3 percent). The percentage changes were not significantly different in these two states.

percentage of the population 5 years and over who spoke a language other than English at home.

The South (2.1 percentage points) and the West (2.4 percentage points) had the greatest gains in the percentage of people who spoke a language other than English at home. The Northeast (1.4 percentage points) and the Midwest (1.0 percentage point) also had gains to the percentages of people who spoke a non-English language.[14] Among the states and the District of Columbia, the percentage-point change ranged from –2.2 percentage points to a change of 4.7 percentage points.

The differences displayed in Table 3b shows that the percentage changes between the C2SS and Census 2000 mirror the percentage changes from Table 3a. While the Northeast, the South, and the West all had significant differences in the number of people who spoke a non-English language, the differences were relatively minor. The percentage-point change in the Midwest was not statistically different.

The 22 states that had significant differences in the number of people who spoke a language other than English at home in Table 3a were the same states that had significant changes in the percentage of their population 5 years and over who spoke a language other than English at home.[15]

[14] The difference in the percentages of people who spoke a language other than English at home was not statistically different between the North (1.4 percentage points) and the Midwest (1.0 percentage point).
[15] South Carolina was the only state that had a significant percentage-point change in the percentage of people 5 years and over who spoke a language other than English but not a significant percentage change in the number of people who spoke a language other than English at home.

Linguistic isolation.

A linguistically isolated household is one in which no person 14 or over speaks English at least "Very well." That is, no person 14 or over speaks only English at home, or speaks another language at home and speaks English "Very well."

A linguistically isolated person is any person living in a linguistically isolated household. All the members of a linguistically isolated household are tabulated as linguistically isolated, including members under 14 years who may speak only English.

The total number of households increased between the 2005 ACS (111 million) and the C2SS (105 million). While most households were not linguistically isolated, the number of households that were linguistically isolated increased for the four major language groups. Linguistically isolated Spanish-speaking households increased from 2.5 million (25.0 percent) in the C2SS to 3.4 million (27.6 percent) in the 2005 ACS. Linguistically isolated households for other Indo-European language-speaker also increased between the C2SS (777,000 households, 15.4 percent) and the 2005 ACS (846,000 households, 16.3 percent).[16]

While the number of linguistically isolated households increased between the C2SS and the 2005 ACS for Asian and Pacific Island languages and for Other languages, the percentage of households that were linguistically isolated did not increase within those two language groups.

[16] The household language is determined by the language spoken by the householder.

As in the case of the other characteristics described in this report, Census 2000 captured more households with individuals who spoke a language other than English at home. The numbers and the percentages of households between the C2SS and Census 2000 were statistically different for households that spoke Spanish, other Indo-European languages, Asian and Pacific Island languages, and Other languages.

The differences between the C2SS and Census 2000 may be due to different data collection methods, different sample frames, from different residence rules, or from the concerted effort of Census 2000 to capture the historically hard-to-capture populations, such as non-English language speakers.

SUMMARY

Survey estimates indicate that the use of a non-English language at home increased and that the English-speaking ability decreased between the 2005 American Community Survey (ACS) and the Census 2000 Supplemental Survey (C2SS). The increase reflected the real-world change where the number of people who spoke non-English languages increased between 2000 and 2005. Comparisons of language use and English-speaking ability between the C2SS and Census 2000 revealed that Census 2000 consistently captured more people with non-English language uses at home. These differences may result from different data collection methods, different sample frames, different residence rules, or from the concerted effort of Census 2000 to capture the historically hard-to-capture populations, such as non-English language speakers.

References

Deborah Griffin, Susan Love, and Sally Obenski, "Can the American Community Survey Replace the Census Long Form?" Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Nashville, TN, May 14-18, 2003.

U.S. Census Bureau, 2003 American Community Survey, "Accuracy of the Data (2005)," Internet document on the source and accuracy of the 2003 ACS data. <http://www.census.gov/acs/www/Downloads/ACS/accuracy2005.pdf>

U.S. Census Bureau, Accuracy of the Data (2000), http://www.census.gov/acs/www/Downloads/ACS/Accuracy00_C2SS.pdf

U.S. Census Bureau, Accuracy of the Data (2005), http://www.census.gov/acs/www/Downloads/ACS/accuracy2005.pdf

U.S. Census Bureau, Census 2000 Frequently Asked Questions. http://www.census.gov/dmd/www/genfaq.htm.

U.S. Census Bureau, *Census 2000 Summary File 3: Technical Documentation.* Washington, DC, 2002.

U.S. Census Bureau, *Design and Methodology: The American Community Survey.* Washington, DC, 2006.

U.S. Census Bureau, *Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey. Report 9: Comparing Social Characteristics with Census 2000.* Washington, DC, 2004.

U.S. Census Bureau, Using the Data: Quality Measure, http://acsweb2.acs.census.gov/acs/www/acs-php/quality_measures_response_2006.php